



Self-Injury and Usage of Social Media Analysis

By
Hayal DARGIN

The Problem

High use of social media has a negative impact on young people which leads them to self harm.

The Solution

To reduce injury rates and suicide attempts there should be strict restrictions on the number of social media apps that teenagers aged between 10 to 19 have access to on their phones and on any other devices.

1. Data Set								
Sex	Race/Ethnicity	Cause of Injury	Disposition	Age Group	Year	Injuries	Popula	
Male	White Non-Hispanice	Self-harm All Injury Causes	All Cases	00-04	2001	0	586	
Male	Black	Self-harm All Injury Causes	All Cases	00-04	2001	0	158	
Male	Hispanic	Self-harm All Injury Causes	All Cases	00-04	2001	0	190	
Male	Other Non-Hispanic	Self-harm All Injury Causes	All Cases	00-04	2001	0	51	
Female	White Non-Hispanice	Self-harm All Injury Causes	All Cases	00-04	2001	0	557	
...		
Male	Other Non-Hispanic	Self-harm All Injury Causes	All Cases	85+	2018	0	11	
Female	White Non-Hispanice	Self-harm All Injury Causes	All Cases	85+	2018	0	334	
Female	Black	Self-harm All Injury Causes	All Cases	85+	2018	0	36	
Female	Hispanic	Self-harm All Injury Causes	All Cases	85+	2018	0	30	
Female	Other Non-Hispanic	Self-harm All Injury Causes	All Cases	85+	2018	0	19	

2. Data Set

Entity	Facebook	Flickr	Friendster	Google Buzz	Google+	Hi5	Instagram	MySpace	Orkut	Pinterest	Reddit	Snapchat	TikTok
Year													
2002	nan	nan	3.000000	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
2003	nan	nan	4.470000	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
2004	nan	3.675135	5.970054	nan	nan	nan	nan	0.980036	4.900180	nan	nan	nan	nan
2005	nan	7.399354	7.459742	nan	nan	9.731610	nan	19.490059	9.865805	nan	nan	nan	nan
2006	nan	14.949270	8.989854	nan	nan	19.932360	nan	54.763260	14.966180	nan	0.248309	nan	nan
2007	nan	29.299875	24.253200	nan	nan	29.533250	nan	69.299875	26.916562	nan	0.488331	nan	nan
2008	100.000000	30.000000	51.008911	nan	nan	55.045618	nan	72.408233	44.357628	nan	1.944940	nan	nan
2009	276.000000	41.834525	28.804331	nan	nan	57.893524	nan	70.133095	47.366905	nan	3.893524	nan	nan
2010	517.750000	54.708063	nan	166.029650	nan	59.953290	nan	68.046710	49.941613	nan	nan	nan	nan
2011	766.000000	66.954600	nan	170.000000	nan	46.610848	nan	46.003536	47.609080	nan	nan	nan	nan
2012	979.750000	79.664888	nan	170.000000	107.319100	nan	nan	nan	45.067022	nan	nan	nan	nan
2013	1170.500000	80.000000	nan	170.000000	205.654700	nan	117.500000	nan	nan	nan	nan	nan	nan
2014	1334.000000	nan	nan	170.000000	254.859015	nan	250.000000	nan	nan	nan	135.786956	nan	nan
2015	1516.750000	nan	nan	170.000000	298.950015	nan	400.000000	nan	nan	nan	163.346676	nan	nan
2016	1753.500000	nan	nan	nan	398.648000	nan	550.000000	nan	nan	143.250000	238.972480	238.648000	nan
2017	2035.750000	nan	nan	nan	495.657000	nan	750.000000	nan	nan	195.000000	297.394200	nan	239.1
2018	2255.250000	nan	nan	nan	430.000000	nan	1000.000000	nan	nan	246.500000	355.000000	nan	500.0
2019	2375.000000	nan	nan	nan	nan	nan	nan	nan	nan	291.000000	nan	nan	nan

Data Wrangling Self Harm

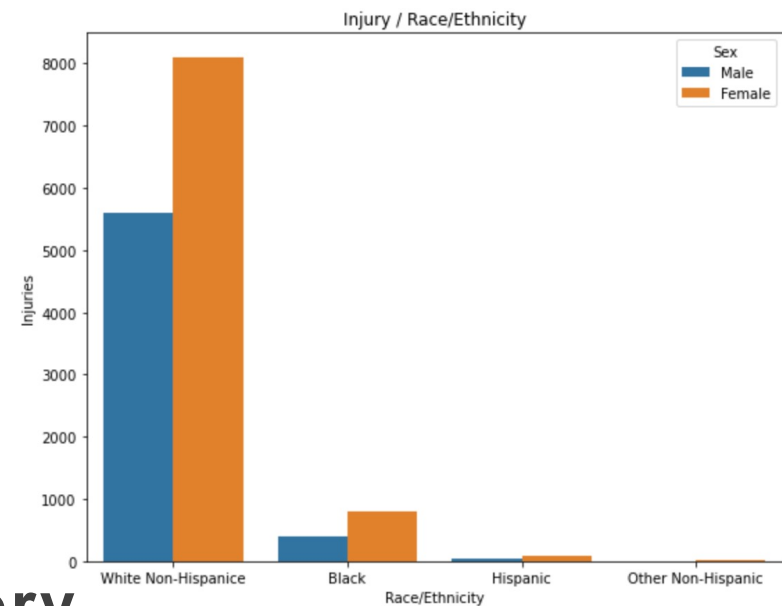
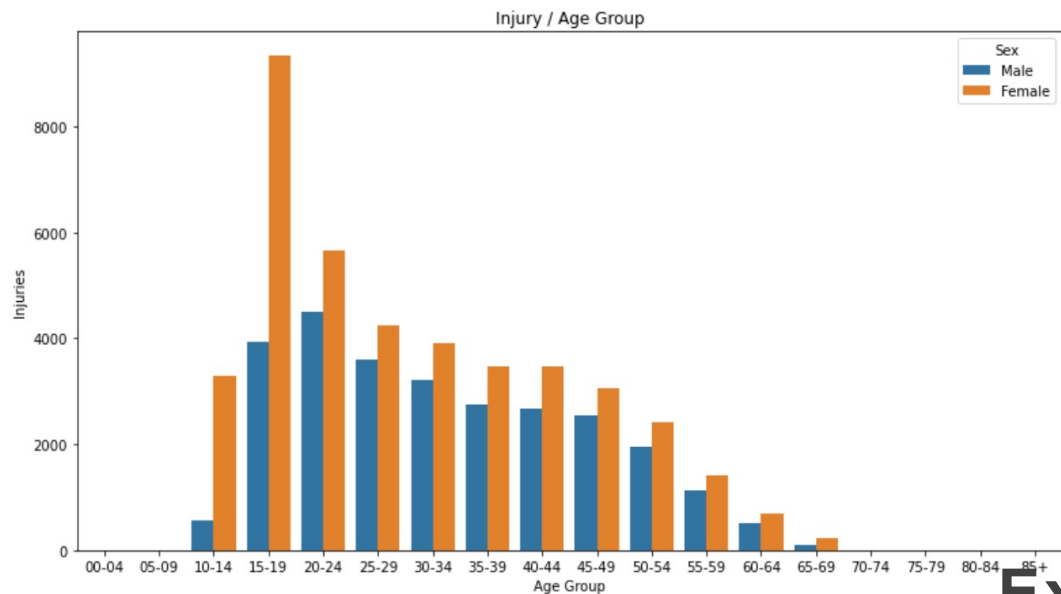
The row data set from CDC (Centers for Disease Control and Prevention) Self Harm contained 5889 rows with 21 columns.

- ❖ Target Variable: Injury
- ❖ Drop redundant columns
- ❖ Drop/Fulfill missing values
- ❖ Correct data types
- ❖ Define unclear values

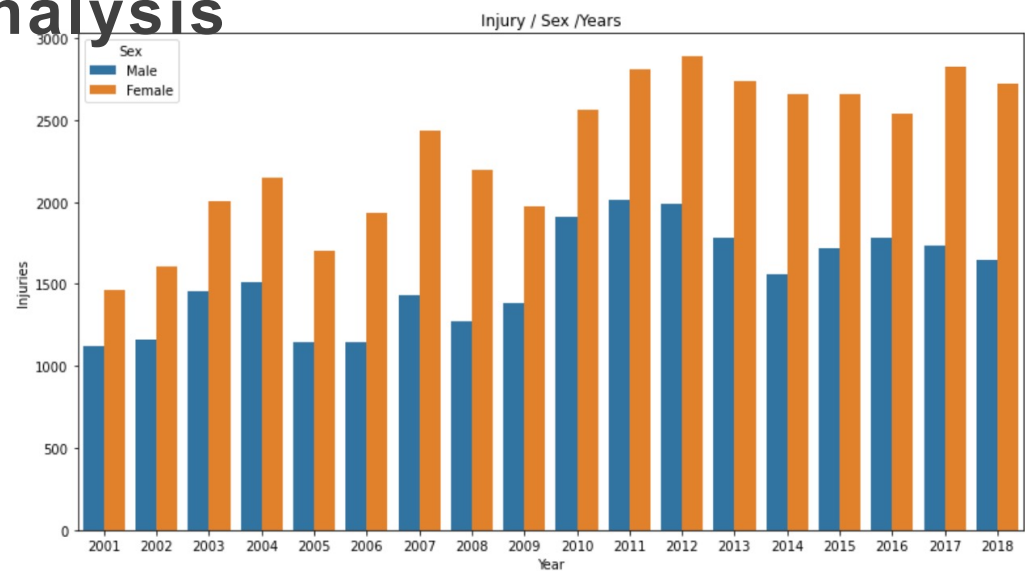
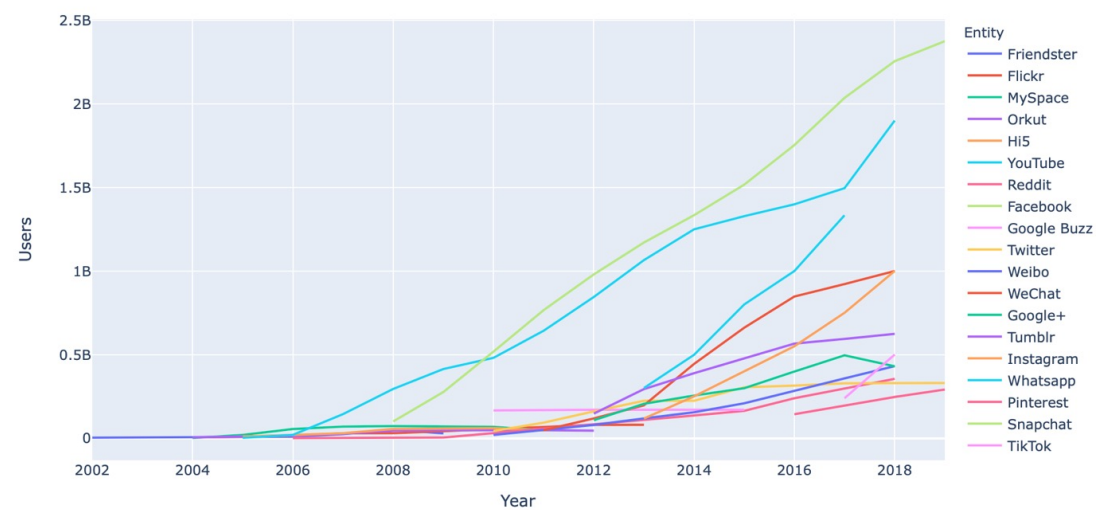
Data Wrangling Social Media

The row data set from Statista Monthly Active Users by My Space contained 142 rows with 3 columns.

- ❖ Target Variable: Users
- ❖ Reshape the data set



Exploratory



Model Selection

- ❖ Linear Regression
- ❖ Random Forest Regressor

Takeaways

- ❖ In Self Harm data set self-injuries of White Non-Hispanic females between the age group 15-19 are more likely to commit self-harm.
- ❖ Random forest regressor was the best model

Future Research

- ❖ Expand the data set
- ❖ Expand modeling
- ❖ Expand to more features in social media data set

Thank You!
